文献与数据学报
第 5 卷　第 4 期　　2023 年 12 月

我国人工智能政策语义、主题特征与演化研究

——基于政策文本的量化分析*

苏丹妮[1]　皮思斯[2]　奉国和[2]
（1. 中山大学信息管理学院，广州　510006；
2. 华南师范大学经济与管理学院，广州　510631）

摘　要：[目的 / 意义] 政策作为特殊因素对我国人工智能行业发展起到指导、规范和约束作用。探究政策出台现状并提出建议，以期促进我国人工智能健康发展。[方法 / 过程] 选取我国最具代表性的国家级和省部级人工智能政策文本共 66 项作为研究对象，通过 BERT-Based 混合 LDA 模型建模和语义量化分析方法，从政策语义、主题特征及其演化对其内含语义和外部属性两方面进行全面探究。[结果 / 结论] 目前我国人工智能政策工具以需求型为主，政策主题以促进领域融合为主，重视基础技术研究和实际成果转化；提出我国需建立人工智能政策统筹规划协调机制，在重视人工智能风险治理的同时，探索实际效益转化的新模式、新路径，并不断深化国际合作和竞争等政策建议。

关键词：人工智能　政策文本分析　主题特征分析　语义量化分析　LDA 模型　BERT-Based 混合模型

分类号：G203；D601

DOI：10.31193/SSAP.J.ISSN.2096-6695.2023.04.02

0 引 言

人工智能正成为引领世界发展的新兴驱动力[1]。随着国家经济社会发展进入新阶段，我国

* 本研究系 2021 年广东省哲学社会科学规划项目“多元动态的科技人才评价方法与实证研究”（项目编号：GD21CTS05）、2020 年广州市科技计划项目（基础与应用基础研究专题）“突发词探测理论、方法与应用研究”（项目编号：202002030384）和 2021 年度华南师范大学哲学社会科学重大培育项目“多元协同的政策文本量化分析方法与实证研究”（项目编号：ZDPY2107）研究成果之一。
[作者简介] 苏丹妮（ORCID：0009-0003-4323-1220），女，硕士研究生，研究方向为数据挖掘、文献与文化遗产保护，Email：2971530142@qq. com（通讯作者）；皮思斯（ORCID：0009-0002-0276-2618），女，本科生，研究方向为数据挖掘，Email：951028255@qq. com；奉国和（ORCID：0000-0002-0774-1544），男，教授，博士，研究方向为数据挖掘、信息计量学，Email：ghfeng@163. com。

人工智能的发展受到重视与推动，更多的人工智能与智能系统研究项目获得国家重点支持，促进人工智能的多领域交叉融合发展，这将催生我国新一轮科技革命和产业变革，成为推进供给侧结构性改革的新动能和建设制造强国的新引擎[2]。然而，处于发展初期的新兴产业往往伴随着市场不经济性和波动性，需要政策的有效正向引导以提供良好的科技发展环境和市场公平进入与竞争机制[3]。加强对人工智能政策文本的分析与理解，有利于优化该领域政策制定，并支持企业对政策变化趋势的预测，促进我国人工智能健康发展。本研究选取国家级及省部级代表性人工智能政策共 66 项作为研究对象，利用 BERT-Based 混合 LDA 模型建模和语义量化分析的方法，对政策语义、主题特征及其演化进行研究，探究政策出台现状与不足并提出相应建议，完善该领域理论研究。

1 文献综述

目前，人工智能发展势头迅猛，作为重要要素协同各领域各行业发展，政策法规在其发展中始终起着方向引领和资源配置的重要作用。然而，针对人工智能政策的理论研究却稍显落后，成果较少。2023 年 4 月 8 日，在 CNKI 检索系统中，以“政策文本分析”为主题词进行全库检索，获得 2786 条结果；以“政策分析”为主题词进行全库检索，检索结果多达 4. 54 万条。而通过专业检索表达式 SU=（SU=‘人工智能’AND SU=‘政策分析’）OR SU=（SU=‘人工智能’AND SU=‘政策文本分析’）在全库范围内进行精确检索后，仅有 55 条检索结果。

现有人工智能政策研究主要从内容量化、文献计量和语义分析三个方面展开。内容量化方面，多以政策工具理论出发，通过对政策文本编码并构建多维框架进行分析。如吕文晶等[4]、毛子骏等[3]基于政策工具和创新价值链搭建二维分析框架，分别将政策文本进行内容分析单元编码，结合 PMC 指数模型对当前中国人工智能产业政策制定的现状与存在的问题进行计量与分析；汤志伟等[5]运用二元 logistic 回归方法，从政策目标、工具和执行三个维度对近年中美两国人工智能产业政策进行对比分析并提出政策建议。文献计量方面，主要针对人工智能政策外部属性进行量化分析。如金双龙等[6]针对人工智能政策的发布数量、时间、效力进行量化分析，为后续政策的比较评价提供数据基础；单晓红等[7]通过量化政策文种、数量和主体属性等特性，针对我国国家级代表性人工智能政策出台前后经济集中发达区域的变化进行分析。语义分析方面，常结合数据挖掘、自然语言处理等方法探究文本内涵意义。如袁野等[8]选择文本挖掘的方法，运用 Python、Gephi 等文本挖掘与图谱分析法对我国人工智能规划进行了量化分析；张涛等[9]通过比较我国各省份人工智能政策文本相似度，得到各地区政策制定的相异之处并对此提出建议；宋伟等[10]对我国人工智能地方性政策进行聚类分析和社会语义网络分析，得到其主题聚类和分布特点。

目前，我国人工智能政策研究处于初级阶段[11]，相关成果较为单薄，研究视角和研究方法也较为类似。部分研究仅针对国家级政策，或仅选取某一省市政策文本进行分析，研究结果缺乏整体代表性，在政策工具类型和效力的框架制定的研究上有待进一步补充和丰富。本研究创新性体现于使用 BERT-Based 混合 LDA 模型对文本主题进行挖掘，在有效把控政策文本 BERT 的语

义信息和LDA的主题信息权重的情况下，发掘潜在的文本主题，更好地探究人工智能政策文本主题演化过程，并从其内含语义和外部属性两方面，对政策语义、主题特征及其演化进行全面探究，选取国家级及省部级近五十万字代表性人工智能政策文本语料作为研究对象，研究范围更为全面，研究对象更具代表性。

2 研究框架与数据来源

2.1 研究框架与方法

参考学者对于政策文本分析的维度[12-14]，政策内含语义指从政策工具视角深入文本内容探究，而政策外部属性包括政策主题词、发文时间与机构等。本研究从政策语义、主题特征及其演化对我国人工智能政策的“内含语义”和“外部属性”两个方面展开分析。内含语义方面，结合政策工具类型和政策效力两个维度进行分析。政策工具类型的结构反映政策制定者的价值偏好和发展模式导向，其量化探究可更全面地了解人工智能政策所内含的价值属性或利益属性[15]。对政策效力的量化体现政策的权威性以及各级政策制定主体对于人工智能领域发展的重视程度。政策工具类型和政策效力二维分析可更加透彻地理解政策决策者的价值导向和未来发展趋势。外部属性方面，核心为政策主题特性与演化分析，通过BERT-Based混合LDA模型进行训练并实现输出，说明关注度较高的政策及其变化情况与政府在引领某一领域发展的内在逻辑[16]，帮助决策者跟踪创新及知识流动[17]。通过“内含语义”和“外部属性”两部分进行分析，将对我国人工智能政策的出台与引领方向有更全面的理解。本研究分析框架如图1所示。

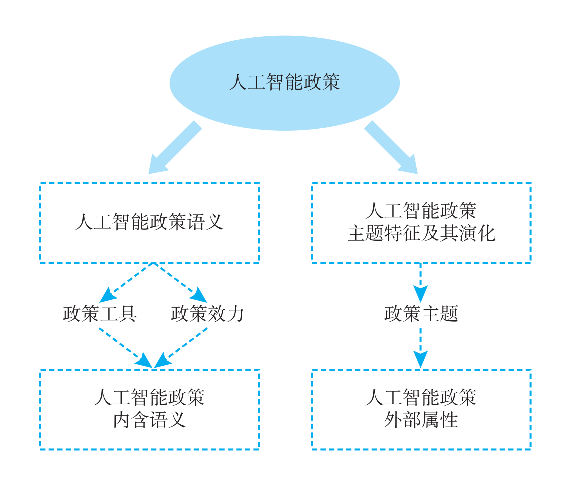


图1 人工智能政策文本量化分析框架

2.2 数据来源与文本选择

考虑数据完整性与检索系统易用性，本研究数据主要来源于“北大法宝”数据库（http://www.pkulaw.cn/），在此基础上辅以检索工业和信息化部、科学技术部、交通运输部、国家发展改革委、国家卫生健康委等国家机构官方网站进行政策文本数据收集。检索过程均以“人工智

能”为检索关键词，进行同义词类的精准搜索。检索时间为 2023 年 4 月 8 日。检索结果显示，与“人工智能”相关的中央法规及各部门政策条例共 100 篇，其中现行有效的为 97 篇。

为保证研究的代表性和时效性，按照以下标准进行政策文本选择：（1）优先选取国家级效力政策文本，以国务院及下属各部委发布的政策文本为主，部分省市级代表性文件为辅；（2）以近年来发布的政策文本为主，具体区间为 2015~2023 年发布的政策文本；（3）选取政策内容与人工智能发展紧密相关的文件，包括直接针对人工智能发展（即标题中含“人工智能”）的文件，及政策文本包含内容关联性较强的文件；（4）筛除已失效的政策文本。最终整理出 66 篇与人工智能相关性最强的政策样本，全文近 50 万字，作为本次研究的文本语料。

3 人工智能政策内含语义分析

根据我国人工智能领域重大政策文件出台时间，可将我国人工智能政策发展分为三个阶段：探索发展阶段（2017 年及以前，以下简称 T1）、国家引导阶段（2018~2019 年，以下简称 T2）和稳定发展阶段（2020 年及以后，以下简称 T3）。初步筛选后，三阶段分别包含政策文本数量 5 篇、19 篇和 42 篇。T1 时期缺乏国家级宏观政策的统筹引导与支持，只有零星行业、部门出台的相关政策，呈现政策量少、分散且内容覆盖范围局限的特点；T2 时期由于国家级人工智能行动计划、发展规划等起步，对人工智能行业发展重视程度提高，政策数量和政策效力显著提升；T3 时期，政策逐步加大对人工智能行业发展干预，行业标准、规范等高效力政策陆续出台，具体实施条例等低效力政策稳步落实。

3.1 政策工具维度

政策工具理论将政策工具从工具作用层面分为供给型、需求型和环境型三类，平衡使用这三种类型的政策工具能保证政策的合理性与科学性[18]。其中，供给型政策工具起到直接推动作用；需求型政策工具通过刺激内需，起拉动作用；环境型政策工具起间接影响作用。供给型、需求型和环境型政策工具对我国人工智能发展作用如图 2 所示。

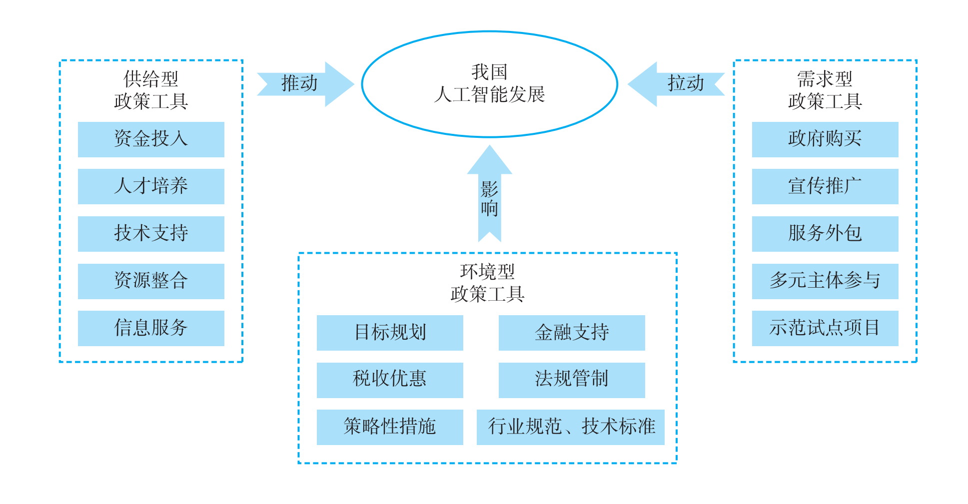


图 2 政策工具与我国人工智能建设作用关系图

供给型政策工具指政府对人工智能产业人才、技术和直接的资金投入支持及信息技术支撑，并为帮助行业发展进行资源整合，是政策支持对于人工智能发展的重要表现。需求型政策工具通过各类条例刺激人工智能市场需求，积极引导试点项目并提供资金扶持，拉动行业快速发展。环境型政策工具监管人工智能建设与发展，创造良好发展环境，间接地对其产生促进作用。值得注意的是，政策类型划分不能单以政策标题作为判断，如《新一代人工智能发展规划》和《促进新一代人工智能产业发展三年行动计划（2018—2020 年）》两部重要指导文件，虽然政策标题仅属于需求性政策，但是内容涵盖供给型、需求型及环境型三方面内容，是顶层的总体规划指导政策。对 T1、T2 和 T3 时期人工智能政策文本内容（包含复合型政策）的工具类型分类统计，得到如图 3 所示。

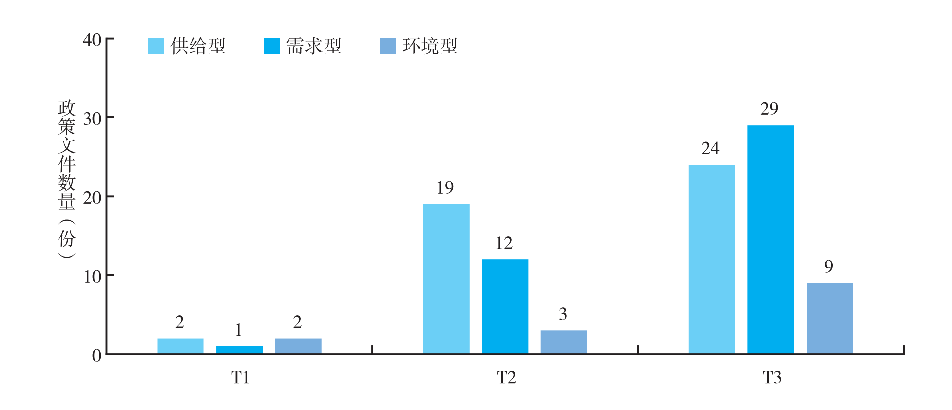


图 3　我国人工智能政策工具类型数量统计

T1 至 T3 三阶段政策工具类型变化特征明显，人工智能政策总体数量与各类型政策工具数量明显增多，可见我国宏观层面对于人工智能领域发展的重视程度不断提升、涉及领域不断拓展。分类型而言，我国从以供给型人工智能政策工具为主转变为以需求型政策工具为主。T1 和 T2 处于初步探索和国家引导阶段，依靠发布供给型政策进行推动，进入 T3 时期后，政府相应地增加需求型政策发布，激发我国人工智能与多领域间交叉融合发展的内化活力，调动协同发展活力。在不断正向发展的大环境下，环境型政策工具也同步增加，为未来的良好发展奠定环境支持基础。

3.2　政策效力维度

根据《党政机关公文处理工作条例》（中办发〔2012〕14 号）[19]，结合我国人工智能政策文本特点，本研究将政策效力量化为五类，结果如表 1 所示。

表 1　人工智能政策效力量化标准

指标	级别（得分）	评判标准
政策力度	5	国家级、省市级地方性法规和司法解释
	4	国家级发展规划 / 行动计划（建设指南）、政府及各部门颁布的规章与规范性文件（规范 / 指引 / 标准）
	3	政府及各部门颁布的条例、规定、决定、实施方案（办法 / 细则）
	2	政府及各部门的工作计划（工作方案）、意见、办法、暂行条例 / 试行规范、批复
	1	政府及各部门的通知、函、通告（公告）

依据上表等级分值并结合前文阶段划分，T1、T2 和 T3 时期我国人工智能政策效力量化得分统计分别为 11 分、29 分和 63 分。2017 年为 T1 至 T2 的转折点，是我国人工智能发展的关键一年，自 2017 年人工智能首次被纳入国家发展规划起，政策效力量化得分开始明显增加。三阶段政策效力量化得分呈现总体稳步增长状况，这意味着国家对人工智能行业发展重视程度上升，整体呈现良好发展态势。

我国人工智能相关政策颁发单位涵盖国家级及省部级各政府单位，覆盖领域广泛。本研究时间范围内，国家层面多主体合作制定政策共 12 项，其中国家发展改革委及工业和信息化部参与较多，联合主体大部分为“省市人民政府 + 信息化委员会 +X”的组合模式。由涉及政策出台的政府部门领域铺开、数量增多、力度增强的趋势可见，人工智能产业展现出多领域融合的积极尝试，实现“人工智能 + 旧产业 = 新产业”的突破。政策制定总体以国务院及其下辖部门为主，其余国家级政府组织及省市部门做好紧密配合，补充完善人工智能政策发布单位总体架构，做到上层规划、下层部署、底层落实，实现等级与职能互补。

3.3 政策工具 – 政策效力二维量化分析

抽取 2017~2021 年间我国人工智能政策工具 – 政策效力二维量化数据如图 4 呈现。2017 年和 2019 年作为 T1 阶段与 T2 阶段的转折年份，阶段特色明显。2017 年国务院印发《新一代人工智能发展规划》和工业和信息化部印发《促进新一代人工智能产业发展三年行动计划（2018—2020 年）》两大标志性政策，是我国人工智能领域首度由国家层面做出政策指示，标志着我国人工智能发展进入 T2 阶段。2017 年之后，供给型政策数量明显增加，各级政府响应国家指示积极推进人工智能发展，政策效力也随之提升。2019 年之后，国家层面相继出台《国家新一代人工智能创新发展试验区建设工作指引》和《国家新一代人工智能标准体系建设指南》这两部属于供给型和需求型的政策类型文件，标志着我国人工智能领域由国家引导阶段（T2）进入 T3 阶段。此阶段供给型政策出台数量稍有减少，需求型政策稳步增长并成为主要政策工具类型。而此阶段的政策效力得分进一步跃升，可见各级政府对于人工智能领域发展深度不断推进，政策导向更为积极。

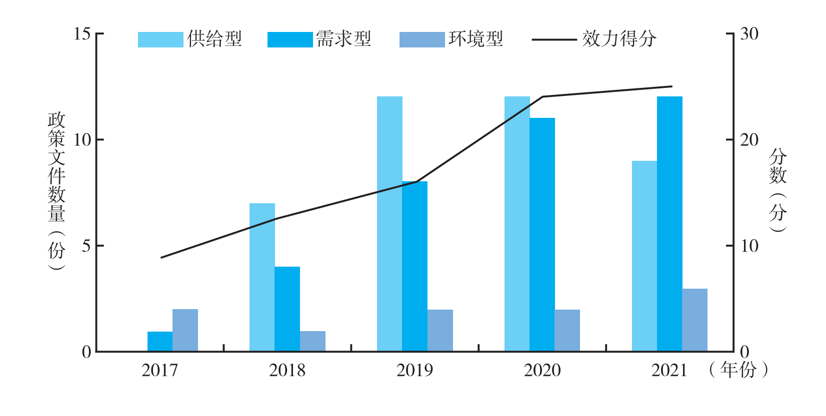


图 4　人工智能政策工具 – 政策效力二维量化数据（节选）

4 人工智能政策外部属性分析

本研究采用 BERT–Based 混合 LDA 模型对我国人工智能政策文本进行外部属性分析，包括以下过程：文本数据预处理和数据集构建、BERT–Based 混合 LDA 模型训练和模型评估及优化，具体过程如图 5 所示。需要注意的是，此处所提的 BERT–Based 混合 LDA 模型是指以 BERT 模型为基础与 LDA 模型混合使用，与 BERT–LDA 模型不同，其最大差异在于 BERT–Based 混合 LDA 模型可根据针对语义信息和主题信息在文本分类中需求的不同自定义权重。BERT–Based 混合 LDA 模型可充分利用 BERT 模型的语义信息和 LDA 的主题信息，有助于提高文本分类效果，特别针对人工智能政策文本研究小规模语料、长篇幅文本以及文本分类体系不一致的特点效果更好。

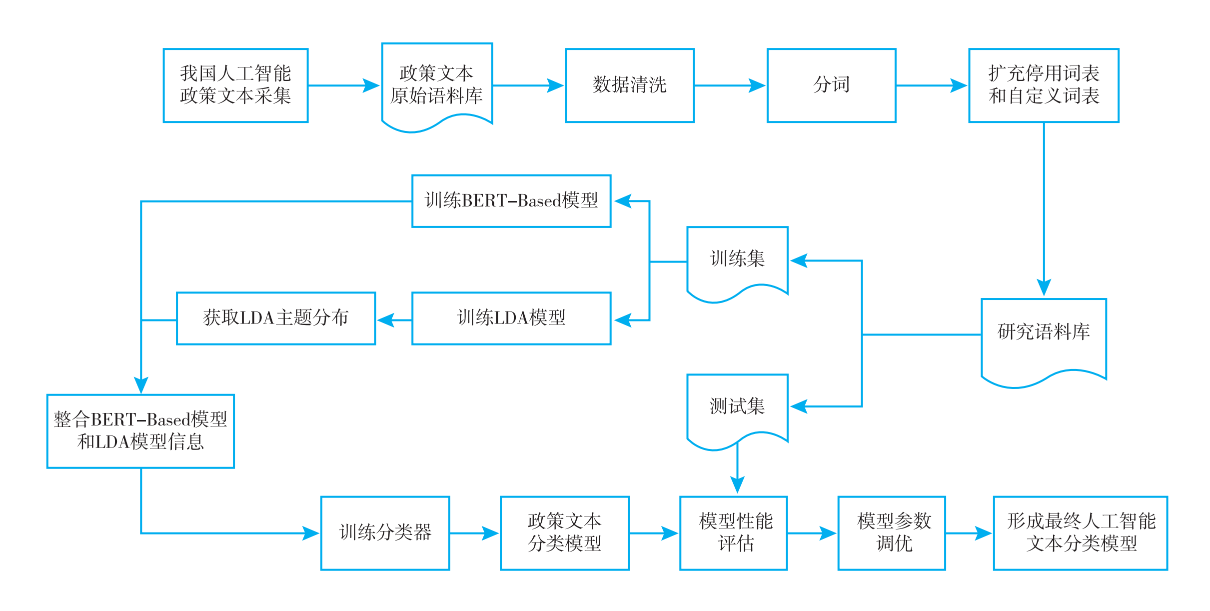


图 5 我国人工智能政策文本基于 BERT–Based 混合 LDA 模型分类过程

4.1 文本向量化表示与模型参数优化

本研究通过加权平均将 BERT–Based 模型与 LDA 混合使用。以下为 BERT–Based 混合 LDA 模型文本向量化表示详细步骤：（1）BERT–Based 文本表示：假设 BERT–Baesd 模型将每个文本表示为一个维度为 d 的向量，表示为 BERT–Baesd_vector，即 BERT–Baesd_vector =［v_1，v_2，…，v_d］。（2）LDA 主题分布：假设 LDA 模型将每个文本表示为一个维度为 n 的主题分布向量，表示为 LDA_distribution，即 LDA_distribution =［p_1，p_2，…，p_n］，其中 p_i 表示文本在第 i 个主题上的概率。（3）文本向量化表示：利用加权平均方式为 BERT–Based 向量和 LDA 的主题分布赋予不同的权重，其中 BERT–Based 向量的权重为 α，给 LDA 的主题分布赋权重 β，则文本向量化表示为 Mixed_vector = α *BERT–Baesd vector + β *LDA_distribution。Mixed_vector 为维度为 d+n 的向量。

在 BERT–Based 混合 LDA 模型中，BERT 的语义信息和 LDA 的主题信息可通过参数调整，不同程度地影响人工智能文本分类。通过对文本语料库不断实验得到的精确率（P 值）、召回率

（R 值）和 F1 值三大评价指标观察，最终得到本研究 BERT-Based 混合 LDA 模型参数最优解 α 为 0.78，β 为 0.22，此时 F1 值、精确率和召回率分别得到 91.12%、91.11% 和 90.92%，即加权平均后的文本向量化表示 Mixed_vector = 0.78*［v_1，v_2，…，v_d］+0.22*［p_1，p_2，…，p_n］。可知，我国人工智能政策文本分类过程中，语义信息贡献度较大，而无监督模型 LDA 占比较小。这也将充分利用 BERT 的语义信息，同时考虑 LDA 的主题信息，使得文本特征更加全面和丰富。

4.2　人工智能政策文本主题特征分析

基于以上实验及模型调参后，利用困惑度确定 T1 至 T3 时期的最佳主题数量，选择困惑度最低的主题数量作为 BERT-Based 混合 LDA 模型主题数量，从图 6 中可以得出 T1 至 T3 时期困惑度最低时对应的主题数量分别为 3、7、14。

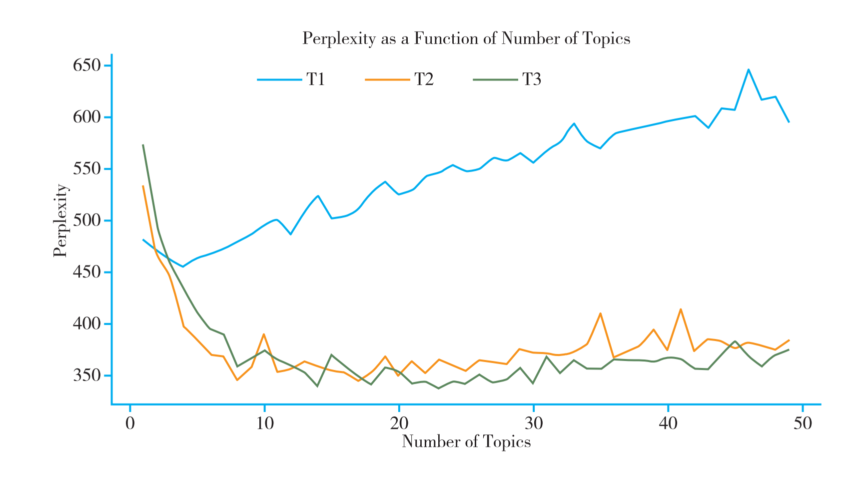


图 6　我国人工智能政策文本各时期主题数 - 困惑度折线图

T1 时期人工智能在国内初露头角，各类人工智能项目涌现，主题内容如表 2 所示。主要为 Topic0 国家对人工智能项目发展的支持，Topic1 关注人工智能产业发展，Topic2 技术发展与平台建设。

表 2　T1 时期主题与主题关键词（部分）

主题序号	前 10 个高频率词及其概率
Topic0	0. 035* 项目 +0. 024* 发展 +0. 014* 工程 +0. 014* 资金 +0. 012* 申请 +0. 012* 改革 +0. 011* 建设 +0. 011* 单位 +0. 009* 国家 +0. 009* 条件
Topic1	0. 038* 人工智能 +0. 033* 智能 +0. 019* 发展 +0. 015* 产品 +0. 014* 技术 +0. 014* 领域 +0. 014* 产业 +0. 012* 行业 +0. 011* 建设 +0. 010* 智能化
Topic2	0. 069* 人工智能 +0. 043* 智能 +0. 022* 技术 +0. 020* 发展 +0. 012* 系统 +0. 011* 研究 +0. 011* 建设 +0. 011* 平台 +0. 010* 重点 +0. 010* 领域

T2 时期人工智能在国内快速发展，融合应用于各行业领域，主题内容如表 3 所示。Topic0 高频关键词体现我国通过政策对人工智能的发展建设进行整体规划；Topic1、Topic6 凸显 T2 时期特色，我国人工智能在快速发展阶段重视基础技术的创新发展，并有意识地进行服务应用建设；Topic2、Topic4 表明此时期我国针对人工智能发展设立研讨会和项目申报支持，促进该领域

知识产权发展，重视人工智能产品发明智力成果保护；Topic3、Topic5 关键词表明此阶段人工智能作为辅助技术，深入物联网领域和医疗领域融合发展。相较于 T1 时期，T2 时期的主题更加具体，人工智能应用领域更加丰富，发展涉及主体更加多元。

表 3 T2 时期主题与主题关键词（部分）

主题序号	前 10 个高频率词及其概率
Topic0	0. 058* 人工智能 +0. 030* 创新 +0. 029* 发展 +0. 021* 建设 +0. 015* 国家 +0. 015* 评估 +0. 015* 政策 +0. 013* 规划 +0. 013* 试验区 +0. 012* 高校
Topic1	0. 023* 智能 +0. 015* 人工智能 +0. 015* 技术 + 0. 013* 产品 +0. 010* 研发 + 0. 010* 服务 +0. 010* 领域 +0. 008* 创新 +0. 008* 产业 + 0. 008* 工业
Topic2	0. 013* 代表 +0. 010* 知识产权 +0. 009* 国际 +0. 009*AI+0. 009* 报名 +0. 007* 人工智能 +0. 007* 机构 +0. 007* 专利 +0. 007* 参会 +0. 007* 研讨会
Topic3	0. 017* 系统 +0. 013* 物联网 +0. 012* 机器人 +0. 012* 工业 +0. 012* 设备 +0. 010* 管理 +0. 010* 工程 +0. 010* 维护 +0. 010* 职业 +0. 009* 技术
Topic4	0. 048* 项目 +0. 042* 申报 +0. 024* 人工智能 +0. 020* 单位 +0. 016* 创新 +0. 013* 产品 +0. 012* 科技 +0. 010* 融合 +0. 010* 行业 +0. 009* 发展
Topic5	0. 042* 医疗器械 +0. 029* 有限公司 +0. 028* 技术 +0. 019* 观察员 +0. 015* 科技 +0. 015* 中心 + 0. 014* 单位 +0. 013* 标准化 +0. 012* 归口 +0. 012* 医疗
Topic6	0. 034* 人工智能 +0. 022* 创新 +0. 021* 技术 +0. 020* 智能 +0. 018* 建设 +0. 015* 平台 +0. 014* 发展 +0. 012* 开放 +0. 012* 领域 +0. 010* 服务

T3 时期主题内容如表 4 所示。其中较为突出的主题有 Topic1、Topic4 医疗行业和知识产权界定，行业标准规范开始出现，利好人工智能领域发展；Topic3、Topic5、Topic7、Topic10 突出人工智能人才培育，充分重视人才主观能动性在人工智能领域发展的推动作用；Topic7、Topic12 通过各类人工智能相关竞赛举办以及针对活动的宣传报道，提升人工智能社会关注度，激励相关领域发展重视和人才培养；Topic0、Topic2、Topic6、Topic8、Topic9、Topic11、Topic13 多次提及人工智能的多场景应用。特别的是，Topic2 中首次关注到人工智能在适老性应用服务中的伦理问题，也是目前全球人工智能发展的焦点问题之一。

表 4 T3 时期主题与主题关键词（部分）

主题序号	前 10 个高频率词及其概率
Topic0	0. 001* 人工智能 +0. 000* 智能 +0. 000* 发展 +0. 000* 场景 +0. 000* 技术 +0. 000* 建设 +0. 000* 领域 +0. 000* 系统 +0. 000* 产品 +0. 000* 标准
Topic1	0. 050* 人工智能 +0. 036* 医疗器械 +0. 021* 产业 +0. 020* 医用 +0. 020* 管理 +0. 020* 知识产权 + 0. 016* 中心 +0. 016* 规范 +0. 016* 界定 +0. 016* 运营
Topic2	0. 051* 人工智能 +0. 029* 老年人 +0. 025* 服务 +0. 023* 风险 +0. 023* 技术 +0. 014* 智能 +0. 014* 产品 +0. 014* 伦理 +0. 012* 发展 +0. 011* 部署
Topic3	0. 087* 人工智能 +0. 045* 发展 +0. 034* 建设 +0. 025* 试验区 +0. 020* 国家 +0. 015* 教育 + 0. 014* 领域 +0. 011* 技术 +0. 011* 科技部 +0. 011* 资源
Topic4	0. 079* 标准 +0. 059* 人工智能 +0. 049* 智能 +0. 033* 技术 +0. 022* 规范 +0. 021* 领域 +0. 017* 服务 +0. 017* 数据 +0. 013* 产品 +0. 011* 模型

续表

主题序号	前 10 个高频率词及其概率
Topic5	0. 058* 教师 +0. 034* 教育 +0. 034* 建设 +0. 029* 人工智能 +0. 027* 教师队伍 +0. 019* 智能 +0.016* 试点 +0. 016* 试点工作 +0. 015* 高校 +0. 013* 技术
Topic6	0. 039* 智能 +0. 034* 项目 +0. 018* 模型 +0. 016* 单位 +0. 015* 系统 +0. 013* 方法 +0. 013* 技术 +0. 012* 人工智能 +0. 012* 场景 +0. 010* 算法
Topic7	0. 042* 人工智能 +0. 034* 司法 +0. 027* 大赛 +0. 025* 技术 +0. 017* 全国 +0. 013* 服务 +0. 012* 人力资源 +0. 011* 比赛 +0. 010* 职业技能 +0. 010* 组委会
Topic8	0. 000* 智能 +0. 000* 人工智能 +0. 000* 技术 +0. 000* 模型 +0. 000* 项目 +0. 000* 单位 +0. 000* 场景 +0. 000* 产品 +0. 000* 系统 +0. 000* 标准
Topic9	0. 147* 场景 +0. 048* 人工智能 +0. 024* 智能 +0. 021* 企业 +0. 017* 技术 +0. 015* 领域 +0. 012* 合作 +0. 010* 数据 +0. 009* 行业 +0. 009* 城市
Topic10	0. 021* 工业 +0. 021* 信息化 +0. 010* 关键部件 +0. 010* 部属 +0. 006* 科技司 +0. 005* 网站 +0. 005* 机关 +0. 005* 全文 +0. 005* 齐备 +0. 005* 高校
Topic11	0. 036* 智能 +0. 031* 人工智能 +0. 022* 场景 +0. 018* 目标 +0. 016* 训练 +0. 015* 技术 +0. 015* 单位 +0. 014* 预期 +0. 014* 系统 +0. 014* 语音
Topic12	0. 057* 广播电视 +0. 031* 网络 +0. 030* 单位 +0. 026* 技术 +0. 024* 项目 +0. 022* 人工智能 +0. 019* 智能 +0. 017* 办公厅 +0. 016* 大赛 +0. 014* 总局
Topic13	0. 061* 单位 +0. 054* 项目 +0. 030* 人工智能 +0. 027* 产品 +0. 020* 医疗器械 +0. 016* 数据 +0. 015* 科技 +0. 013* 智能 +0. 013* 信息化 +0. 011* 环境

4.3 人工智能政策文本主题演化分析

本研究通过计算相邻时期主题之间的余弦相似度继续对我国人工智能政策主题的演化过程进行研究，余弦相似度越大表明两个主题之间的相似度越高。再通过计算相邻时期主题平均相似度作为阈值，认为当相似度超过该阈值时，主题间存在相关性，对结果进行筛选。最终主题演化结果如图 7 所示。

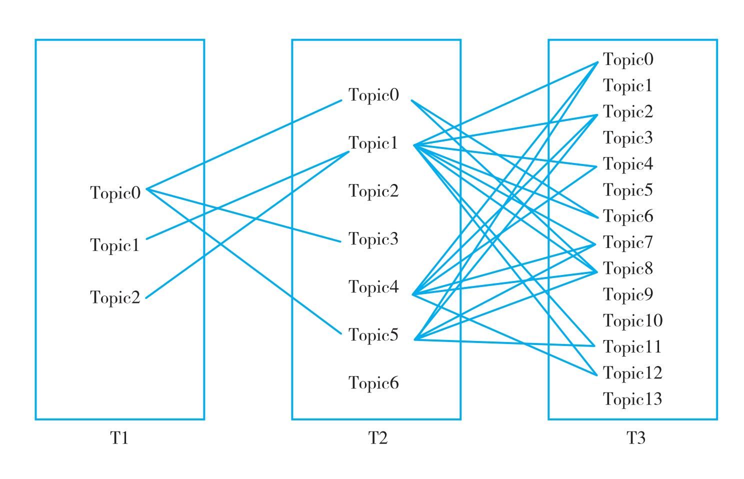


图 7　我国人工智能政策主题演化情况

T1 时期为初步发展时期，人工智能政策文本中均为概括性高、语义内容差距较小的关键词；T2 时期国家积极介入引导与规划人工智能发展，主题词更为明确具体，商业公司作为重要参与主体加入此领域发展，并初步将其应用于物联网、医疗器械等行业生产中；T3 时期人工智能发展迅速，国家政策对其知识产权、行业标准、伦理发展等进行规范，强化人工智能多场景应用、人才队伍建设等。

人工智能产品研发和技术及产业的发展是 T1 时期至 T2 时期相关性最大的主题，并作为重点发展领域延续至 T2 时期。T1 时期重点关注的系统开发、项目建设在 T2 时期得到继承，政策引导与课题评审等主题为 T2 时期新生主题。T2 时期至 T3 时期，相关性最大的主题是人工智能行业标准规范与应用场景，其中产品研发的主题在三个时期中均有提及，说明将人工智能转化为实际应用产品是领域发展重点；政策规划与行业规范在 T2 时期概括性提及，但在 T3 时期呈现在如医疗、物流等多个领域的继承与具体细化，如 T3 时期演化为强调技术板块的算法与模型优化和高校教师队伍培养与试点工作的具体规划等。

5 结论与建议

5.1 研究结论

（1）我国人工智能政策工具以激发内需为主，以塑造环境为辅

我国人工智能发展已进入并且将长期处于稳定发展阶段，通过该领域政策工具变化趋势统计发现，我国人工智能政策已经从依靠供给型政策工具直接推动，转变为由需求型政策发挥主要作用。需求型政策工具通过产业孵化、资金流入、宣传推广等手段，扩张人工智能领域容量，有效激发各行业领域对人工智能的内需活力，进而实现产业升级。而在各阶段发展中，环境型政策起到重要的辅助作用，通过规范行业行为的标准和要求，对人工智能领域发展产生正向影响，塑造适宜的发展环境。

（2）人工智能领域融合逐步扩大，作为全新要素融入现有领域

相关政策出台部门涵盖领域的不断丰富，体现人工智能技术多领域交叉融合范围的扩大。人工智能技术作为一种全新要素，在政策积极推进的加持下赋能现有产业，以“人工智能 + 旧产业 = 新产业”的全新发展模式实现深度产业融合，帮助现有产业，如医疗、养老等领域拓展发展方向，为其提供全新解决方案和优化路径的同时实现自身多维发展，符合技术驱动、数据驱动与需求驱动特征的当下社会发展真实需求。

（3）政策关注人工智能基础技术研究，重视研究成果实际应用转化

基础技术研究是人工智能行业实现可持续发展的关键与动力。国家政策的制定关注实际，在各时期的政策文本中强调人工智能产品研发、技术创新与产业发展等与基础建设相关的主题，体现我国对人工智能基础研究的高度重视和战略布局。同时，以技术的实际转化、产品研发等作为关键词的突出主题，多次出现在人工智能发展各阶段政策文本中。

5.2 尚待完善之处

（1）政策制定和实施统筹协调能力有待提升，需要更多实际数据支撑

各级别、各地区政策制定与实施若缺乏统筹协调机制可能导致政策目标不一致、政策措施不

协调、政策资源不共享等问题出现。由新旧高频关键词研究中可知，国家级政策文本的高频词往往需要在下一阶段才缓慢成为省市级人工智能政策文本关键词，响应存在一定滞后性，存在宏观政策中旧关键词成为下一阶段省市级高频关键词的现象。中央和地方各级人工智能政策制定与实施尚未完全实现统筹规划，各地区政策制定存在一定政策重叠、冲突、缺失，甚至相互抵消等问题，有待形成整体推进的有力格局。此外，各地区政策因实际数据支撑不充分，在准确把握不同地区人工智能发展的现状、趋势和需求方面有待提高。

（2）人工智能领域危机防范意识较为薄弱，风险应对能力有待提升

风险评估和监管机制的政策缺失可能导致难以及时识别和应对人工智能技术的不良影响和潜在危害，无法最大程度地保障个人隐私、数据安全、知识产权等合法权益，避免偏见、歧视、滥用等伦理问题。在 T1 至 T3 阶段，伦理、安全、危机预警等相关词语均未出现在人工智能政策高频词列表当中，人工智能政策需提高对伦理道德、法律、安全等方面的重视，为有效防范和化解人工智能带来的风险和挑战做好准备。

（3）国际合作和竞争战略布局能力有待提升，需提高现有资源利用率

我国现有人工智能政策的制定和实施主要聚焦于国内各地区发展，在国际层面的合作和竞争战略布局能力有待提升，需要进一步整合利用国际平台资源以及更好地吸取发达国家和地区的人工智能政策制定发展的相关经验。

5.3 政策建议

（1）统筹规划人工智能政策制定和实施，优化政策效力结构

横向上，建立健全跨部门、跨地区的政策协调机制，统筹规划全局人工智能政策制定和实施，政策工具类型的选择以实际需求和情况为准，构建符合我国国情的人工智能政策治理框架。纵向上，需要加快人工智能领域法规建设，填补相关领域的政策空白。各级各地区可参考《新一代人工智能发展规划》等国家战略文件，结合我国人工智能发展现状和未来趋势，出台相关的法律法规，为人工智能技术研发、应用和监管提供更多法律依据。

（2）推动政策制定与实体经济深度融合，探索新模式、新路径

人工智能技术最终需要通过与实体经济深度融合转化为实际效益。在此过程中，需要充分发挥数据的基础资源作用，利用政策制定和实施驱动构建数据驱动、跨界融合、共创分享的智能经济形态，加强数据开放共享和流通利用，激发数据价值和创新活力。并需要以实体经济需求为导向，深入挖掘人工智能在各类应用场景和落地方案当中的作用，形成技术和产业的良性循环，推动人工智能与制造、交通、医疗、农业等各领域融合日益深入，持续推动质量、效率及动力变革。

（3）重视人工智能风险治理，建立健全监督机制和标准体系

在政策制定的过程中加强人工智能相关法律、伦理、社会问题等研究，为整个社会建构人工智能时代伦理体系和司法体系的基础理论和价值目标，设计适用于不同群体和应用场景的伦理规范和评判指标，提供人工智能风险治理的深层理论资源。随着人工智能技术与各类领域融合发展，需要重视过程中涉及的风险和挑战，建立政府、公众、媒体、企业和第三方评估机构等多元互动的生态行政管理系统，打造有效约束和容错纠错的管理机制。同时，提升社会反馈响应效率，及时修订完善不适应人工智能发展需要的政策措施。

（4）深化国际合作和竞争，积极参与全球人工智能领域治理

加强国际交流与合作，深入了解各国和地区的人工智能政策制定及战略布局，借鉴各国人工智能政策制定、技术创新、产业应用、人才培养等方面的经验。关注国际组织和论坛，如联合国、G20、OECD 等在人工智能治理方面的倡议和行动。积极参与全球人工智能治理的制定和实施，提升我国在全球人工智能治理中的话语权和影响力，推动构建符合我国利益和价值的人工智能规则体系。

【参考文献】

[1] 许晔. 下一代人工智能：引领世界发展的新兴驱动力 [J]. 人民论坛・学术前沿，2017（20）：14–23.

[2] 工业和信息化部. 促进新一代人工智能产业发展三年行动计划（2018—2020年）[EB/OL].（2017-12-13）[2023-5-10]. http：//www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5960820/content.html.

[3] 毛子骏，梅宏. 政策工具视角下的国内外人工智能政策比较分析 [J]. 情报杂志，2020，39（4）：74–81，59.

[4] 吕文晶，陈劲，刘进. 政策工具视角的中国人工智能产业政策量化分析 [J]. 科学学研究，2019，37（10）：1765–1774.

[5] 汤志伟，雷鸿竹，周维. 中美人工智能产业政策的比较研究——基于目标、工具与执行的内容分析 [J]. 情报杂志，2019，38（10）：73–80.

[6] 金双龙，隆云滔，陈立松，等. 基于文本分析的区域人工智能产业政策研究 [J]. 改革与战略，2020，36（3）：44–53.

[7] 单晓红，何强，刘晓燕，等. "政策属性—政策结构" 框架下人工智能产业政策区域比较研究 [J]. 情报理论与实践，2021，44（3）：194–202.

[8] 袁野，于敏敏，陶于祥，等. 基于文本挖掘的我国人工智能产业政策量化研究 [J]. 中国电子科学研究院学报，2018，13（6）：663–668.

[9] 张涛，马海群. 基于文本相似度计算的我国人工智能政策比较研究 [J]. 情报杂志，2021，40（1）：39–47，24.

[10] 宋伟，夏辉. 地方政府人工智能产业政策文本量化研究 [J]. 科技管理研究，2019，39（10）：192–199.

[11] 章小童，李月琳. 人工智能政策与规划的主题结构：基于主题词共现网络分析 [J]. 情报资料工作，2019，40（4）：44–55.

[12] 李明，曹海军. 中国央地政府人工智能政策比较研究——一个三维分析框架 [J]. 情报杂志，2020，39（6）：96–103，53.

[13] 吴宾，杨一民，娄成武. 基于文献计量与内容分析的政策文献综合量化研究：以中国海洋工程装备制造业政策为例 [J]. 情报杂志，2017，36（8）：131–137.

[14] 叶选挺，李明华. 中国产业政策差异的文献量化研究：以半导体照明产业为例 [J]. 公共管理学报，2015，12（2）：145–152，159–160.

[15] Rothwell R, Zegveld W. Reindusdalization and technology [M]. London: Logman Group Limited, 1985: 83–104.

[16] 曹蓉，刘彦芝，王铮. 中国慈善政策合作网络与主题热点演化研究——基于SNA和LDA的大数据分析 [J]. 社会保障研究，2023（1）：41–52.

[17] Chen B, Tsutsui S, Ding Y, et al. Understanding the topic evolution in a scientific domain: an exploratory study

for the field of information retrieval［J］. Journal of Informetrics, 2017, 11(4): 1175-1189.

［18］Rothwell R, Zegveld W. Industrial innovation and public policy: preparing for the 1980s and 1990s［M］. London: Frances Printer, 1981.

［19］中共中央办公厅、国务院办公厅. 党政机关公文处理工作条例［Z］. 2013-2-22.

Study on Semantics, Thematic Features and Evolution of Artificial Intelligence Policies in China: Quantitative Analysis Based on Policy Texts

Su Danni[1] Pi Sisi[2] Feng Guohe[2]

(1. School of Information Management,Sun Yat-sen University, Guangzhou 510006, China;
2. School of Economics and Management,South China Normal University, Guangzhou 510631, China)

Abstract:［**Purpose/significance**］ Policy as a special factor plays a guiding, regulating and restraining role in the development of China's AI industry. This paper explores the current status of policy introduction and makes recommendations with a view to promoting the healthy development of AI in China.［**Method/process**］China's most representative 66 national and provincial AI policy texts are selected as research objects,and through BERT-Based hybrid LDA model modelling and semantic quantitative analysis methods, both their internal semantics and external attributes are comprehensively explored in terms of policy semantics, thematic features and their evolution.［**Result/conclusion**］At present,China's AI policy tools are mainly demand-based, and the policy theme is to promote domain integration, emphasizing basic technology research and actual results transformation. By analyzing the policy deficiencies, it is proposed that China needs to establish a coordinating mechanism for the overall planning of AI policies,explore new modes and paths for the transformation of practical benefits while paying attention to the governance of AI risks, and deepen the international cooperation and competition, among other policy recommendations.

Keywords: Artificial intelligence; Policy text analysis; Topic feature analysis; Semantic quantitative analysis; LDA modeling; BERT-Based hybrid modeling

（本文责编：孔青青）